Prof. Philip Koopman

Carnegie Mellon University

Key Ideas: UL 4600 Safety Standard for Autonomous Vehicles

July 2022

https://safeautonomy.blogspot.com/

Overview

Carnegie Mellon University

- UL 4600 standard for AV safety cases
  - Fully autonomous vehicles
  - Issued April 2020

- Key 4600 ideas:
  - System-level safety case provides direction
  - Vehicle as well as infrastructure and lifecycle processes all matter
  - Safety metrics used for feedback loops
  - Third party component interface protects proprietary info
  - 4600 helps you know that you've done enough work on safety

Goal Based Approach

Carnegie Mellon University

- Traditional safety standards are prescriptive
  - "Here is how to do safety" (process, work products)
    - ISO 26262, ISO/PAS 21448, IEC 61508, MIL-STD 882, etc.

- UL 4600 is goal based
  - "Here is what a safety case should address"
    - Do NOT prescribe any particular engineering approach
      - » Use other safety standards within the safety case context
  - Standard for how to assess a safety case
    - Minimum coverage requirement (what goes in the safety case?)
    - Properties of a well-formed safety case
    - Objective assessment criteria

Example 4600 Clause

Carnegie Mellon University

12.3.1 V&V shall provide acceptable coverage of safety related faults associated with the design phase.

12.3.1.1 MANDATORY:

a) Systematic design defects

b) Design consideration of faults, corruption, data loss, and integrity loss in sensor data

c) Requirement gaps/omissions and requirement defects

d) Response to violation of requirement assumptions

EXAMPLE: Response to exceptional operational environment

e) Identification and description of the intended ODD

f) Acceptable mitigation of aspects of the defined fault model for each component and other aspect of the item

12.3.1.2 REQUIRED:

a) Maintenance procedure definitions

NOTE: While maintenance occurs during the lifecycle, the definition of procedures needs to correspond to design requirements and assumptions made in design regarding maintenance.

b) Operational procedure definitions (including startup and shutdown) and operational modes

c) Faults, corruption, data loss, and integrity loss in data from external sources

d) Faults and failures associated with exceptional conditions that impair risk reduction functionality

e) Hardware and software errata and other third-party component design defects

f) Other faults in safety related functions, component designs, and other designed properties

12.3.1.3 HIGHLY RECOMMENDED – N/A

12.3.1.4 RECOMMENDED – N/A

12.3.1.5 CONFORMANCE:

Conformance is checked via inspection of design and V&V evidence.

Flexible Approaches

Carnegie Mellon University

6.4.1 Each identified hazard shall be given a criticality level and assigned an initial risk assuming the absence of mitigation.

6.4.1.1 MANDATORY:

a) Hazard Log records criticality level and initial risk for each hazard

6.4.1.2 REQUIRED:

a) Use of at least one of the following risk evaluation approaches:

1) Risk table
2) Risk equation (weighted probability times severity)
3) Fault Tree Analysis (FTA)
4) Event Tree Analysis (ETA)
5) Preliminary Item Safety Assessment (PSSA)
6) Hazard Analysis and Risk Assessment (HARA)
7) Bowtie diagram
8) System-Theoretic Accident Model and Processes (STAMP)
9) Field engineering feedback
10) Other relevant risk evaluation approaches

b) Use of integrity level and related techniques

EXAMPLES: Integrity level and related techniques from ISO 26262, IEC 61508; development assurance level from DO-178

…

6.4.1.3 HIGHLY RECOMMENDED:

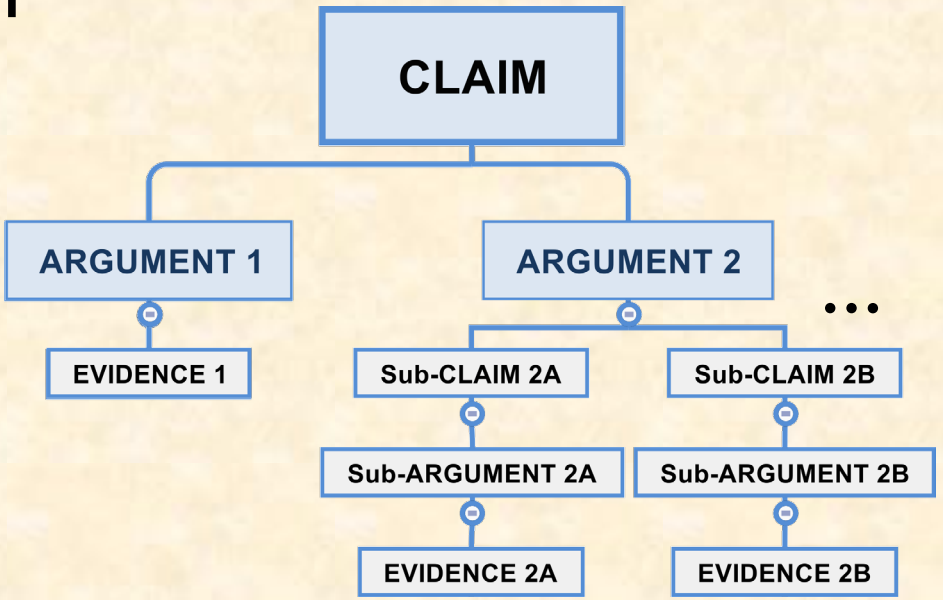
a) Use of integrity levels defined in an accepted domain-relevant functional safety standard

NOTE: It might not be practical to use such integrity levels for all aspects of an autonomous systems, but it is highly recommended to do so to the extent reasonable.

Safety Case

Carnegie Mellon University

- Claim – a property of the system
  - "System avoids pedestrians"
- Argument – why this is true
  - "Detect & maneuver to avoid"
- Evidence – supports argument
  - Tests, analysis, simulations, …
- Sub-claims/arguments address complexity
  - "Detects pedestrians" // evidence
  - "Maneuvers around detected pedestrians" // evidence
  - "Stops if can't maneuver" // evidence



4600 Safety Case Scope

Carnegie Mellon University

- Everything needed to independently assess safety
  - Hazards and mitigation approaches
  - Claims traced: arguments to evidence

- Scope includes:
  - Technology: HW/SW, machine learning, tools, …
  - Lifecycle: deployment, operation, incidents, maintenance, …
  - Infrastructure: vehicle, roads, data networks, cloud computing, …
  - Road users: pedestrians, light mobility, emergency responders, …
  - Environment: Operational Design Domain (ODD) definition
  - … and more …

Example ODD Prompts (§8.2.2)

Carnegie Mellon University

- Behavioral rules
  - EXAMPLES: Traffic laws, vehicle path conflict resolution priority, local customs, justifiable rule breaking for safety
- Compliance strategy of traffic rules and regulations
  - EXAMPLE: Enumeration of applicable traffic regulations and corresponding ego vehicle behavioral constraints

https://bit.ly/2IKlZJ9

- Vulnerable populations including number, density, and types
  - EXAMPLES: Pedestrians, motorcycles, bikes, scooters, other vulnerable road users, other road users
- Special road user rules, if applicable
  - EXAMPLES: Bicycles, motorcycles, lane splitting, interacting with construction vehicles, oversize vehicles, snowplows, sand/salt trucks, emergency response vehicles, street sweepers, horse-drawn vehicles
- Seasonal effects
  - EXAMPLES: Foliage changes (e. g., leaves (dis) appearing), sun angle changes, seasonal behavioral patterns (e. g., summer beach traffic), seasonally-linked events (Oktoberfest, regatta crowds, fireworks gatherings, air shows)

SPI Metrics

- Safety Performance Indicator (SPI)
  - Like a KPI, but specific to safety
  - Provides metrics on safety case validity

- SPI measures:
  - Behavior metrics for safety-related behaviors
    - E.g.: Acceptable violation rate of standoff to pedestrians
  - Assumption validity within safety case
    - E.g.: Tolerates gaps of up to X meters in lane markings
    - E.g.: Correlated camera and lidar false negative rate
  - Any other metrics that validate safety case

Feedback Loops

Carnegie Mellon University

- Rather than assume perfection…

  … manage & improve imperfections

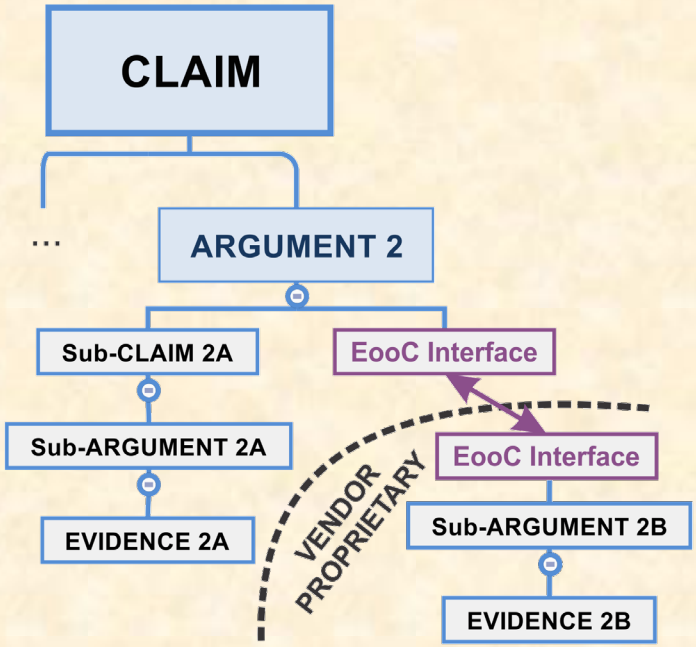
  - Feedback data incorporated in safety case
  - Convert “unknowns” into “knowns” over time

- Feedback loops for continuous improvement
  - Implementation faults
  - Design faults
  - Gaps in simulations, analysis tools, …
  - Gaps in Operational Design Domain
  - Gaps in machine learning training data

[Edge Case Research]

Elements out of Context (EooC)

- Reused or 3rd party system "component"
  - Similar in spirit to ISO 26262 SEooC
  - Hardware, software, sensor, map data, …
- EooC has a safety case fragment
  - Vendor need not expose that safety case
  - Instead, provides an interface containing:
    - Properties & characteristics
    - Assumptions that system must honor
    - Fault model used for assessment
    - 4600 clause coverage (might be partial)
    - Assessment report



Complementing Other Standards

Carnegie Mellon University

- ISO 26262, MIL-STD 882, etc.: potential starting points
  - Still useful where applicable
- ISO/PAS 21448 etc. for scenarios
  - Design and validation process framework
  - SaFAD and emerging standards
- 4600 has #DidYouThinkofThat? lists
  - Initial safety case coverage
  - Learn from experience: yours; others
  - Objective assessment criteria for safety case

Other Key Points

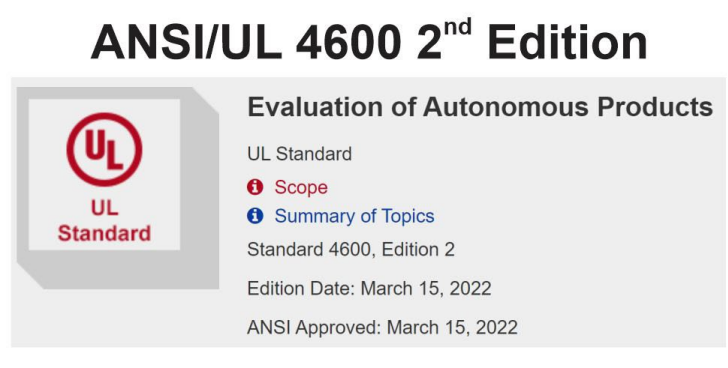
Carnegie Mellon University

- Self-certification is permitted
  - Internal assessor permitted; no external “certificate” requirement
- Only necessary technical mitigations required
  - “Does not apply to this system” and “Outside ODD” are OK
  - Can use non-technical mitigations
- Underwriters Laboratories is a non-profit SDO
  - Voting committee (STP) has diverse representation
  - Continuous Maintenance process provides timely updates
- Does 4600 conflict with ISO 26262 or ISO/PAS 21448?
  - No
- What if you can’t afford to buy a copy?
  - Issued standard is free to browse (“digital view”) on-line in its entirety: https://www.shopulstandards.com/ProductDetail.aspx?productid=UL4600

UL 4600 Version 2

Carnegie Mellon University

- Issued March 15, 2022
- Assessment terminology & roles:
  - Self-assessment
    - Development team vets safety case
  - Independent assessment
    - Scope includes independent technical substance of safety case
- Safety case terminology and structure
  - Significant improvements; same ideas and intent as version 1
- Terminology
  - Improved alignment with other standards
- Other improvements per stakeholder feedback



UL 4600 Version 3 – In Progress for 2022

Carnegie Mellon University

- Primary goal: specific coverage of heavy trucks
  - Expands scope, but no fundamental change was required
- Revised safety case framework for autonomous trucking
  - Adds concept of platoon (coordinated vehicles with a safety buffer)
  - Various related added prompts (e.g., hazardous materials)
- Revised to add examples specific to autonomous trucking
  - Cargo loading/unloading operations
  - Communication with trailing platoon vehicles
- Other improvements
  - Added a preferred Safety Performance Indicator approach
  - Emergency responder terminology

Review of Key Ideas

Carnegie Mellon University

- System-level safety case provides direction
  - Highlights gaps in evidence and arguments
- Vehicle, infrastructure, and lifecycle processes all matter
  - If safety case depends upon it, that makes it safety related
- Metrics combine with feedback loops
  - Operational feedback will be essential for practical safety
- Third party component interface to protect proprietary info
  - EooC interface permits separate component assessment
- 4600 helps you know that you've done enough safety work
  - Robust prompts and pitfalls capture best practice/lessons learned